

What's different about 2020 Census data?

Challenges posed by Census Bureau's new disclosure avoidance system

- Beth Jarosz, Senior Research Associate, Population Reference Bureau
- David Howell, State Demographer, Research and Analysis
- Eric Sandberg, Demographer, Research and Analysis
- Moderator: Liz Brooks, Research Analyst, Research and Analysis



**ALASKA DEPARTMENT OF LABOR
& WORKFORCE DEVELOPMENT**

Sign up to our
listserv to receive
updates.

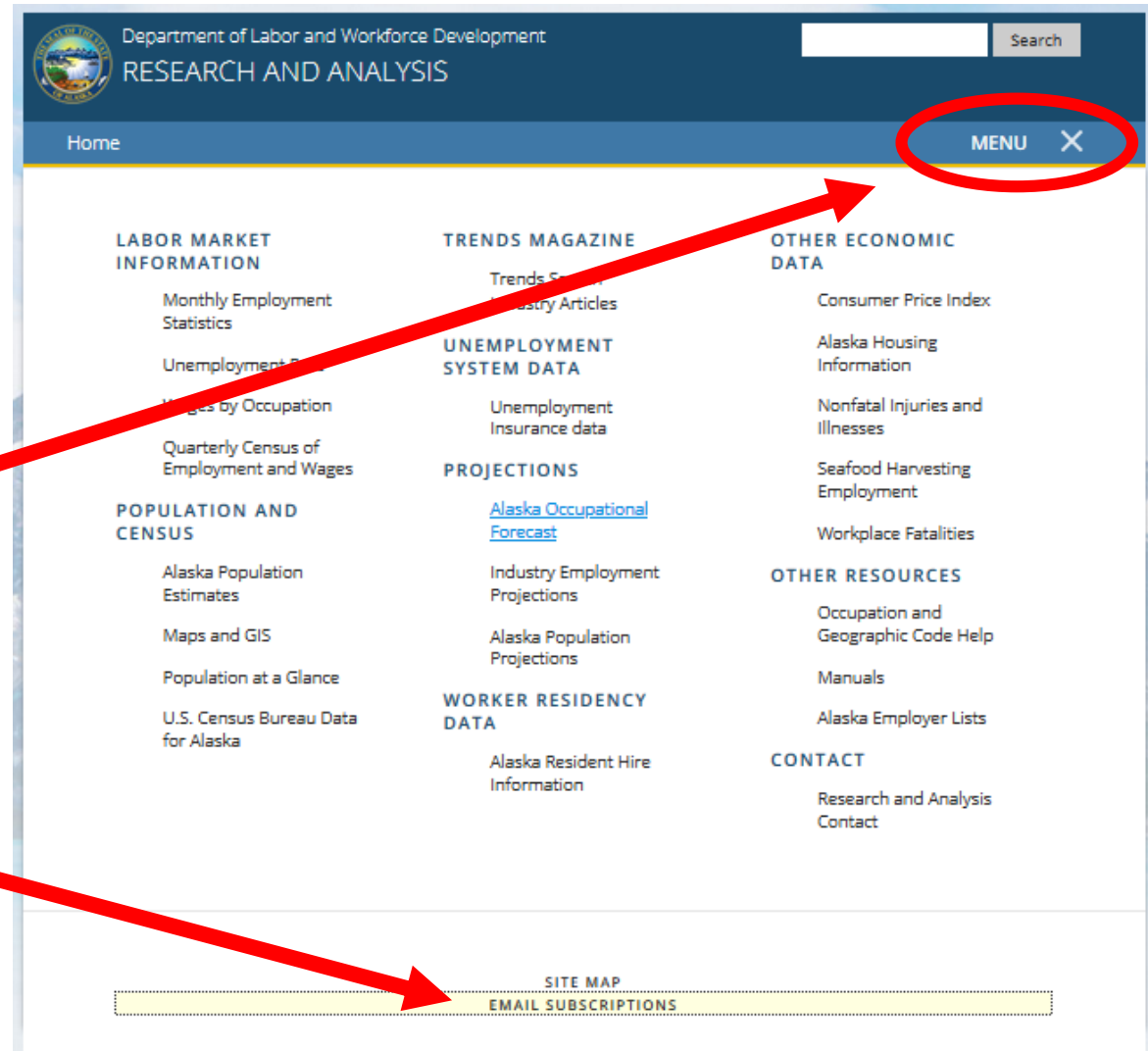
laborstats.alaska.gov

Open MENU then

EMAIL
SUBSCRIPTIONS at
bottom



**ALASKA DEPARTMENT OF LABOR
& WORKFORCE DEVELOPMENT**



PRB

INFORM
EMPOWER
ADVANCE

Disclosure Avoidance and the 2020 Census

Beth Jarosz
@DataGeekB

December 2020

POPULATION REFERENCE BUREAU | www.prb.org

- Why is data privacy an issue now?
- What is the Disclosure Avoidance System (DAS)?
- What does DAS mean for the data?

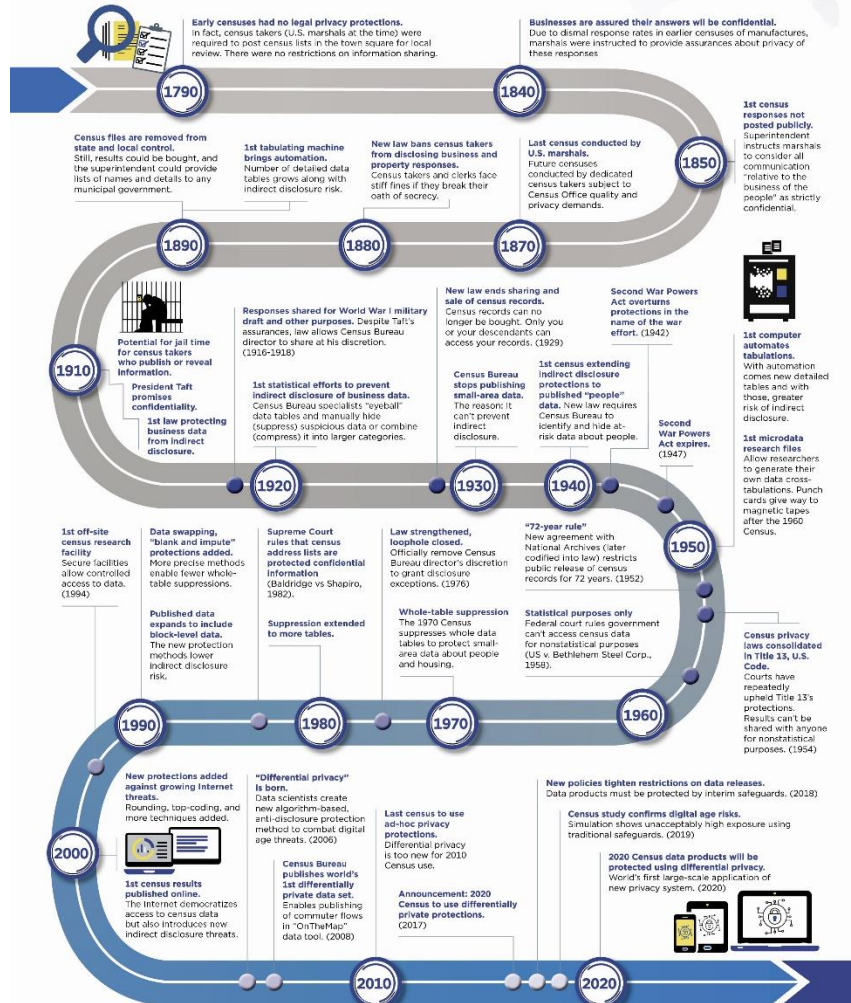
A HISTORY OF CENSUS PRIVACY PROTECTIONS

Today's law is clear: The Census Bureau must keep responses completely confidential. It cannot release identifiable information about an individual, household or business to anyone, including other government or law enforcement agencies.

It wasn't always that way. Public attitudes on privacy have changed since the first census in 1790. Early laws and policies focused on preventing direct disclosure of personal information. Later, laws and policies addressed the growing threat of indirect disclosure—the risk that someone might be able to figure out the identity of a person or business just by analyzing the statistics we publish.

Twenty-first century privacy threats—faster and more powerful computers, new data science, and exponential growth in personal data available online—demand new safeguards to protect against indirect disclosure.

See how the laws and protections have changed from 1790 to the 2020 Census—the first census to use advanced disclosure protections based on the new data science known as “differential privacy.”



U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

To learn more search "Disclosure Avoidance" at [census.gov](https://www.census.gov).

Source

https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

USC Title 13 - Protection of Confidential Information

§9. Information as confidential; exception

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998 or section 2(f) of the Census of Agriculture Act of 1997—

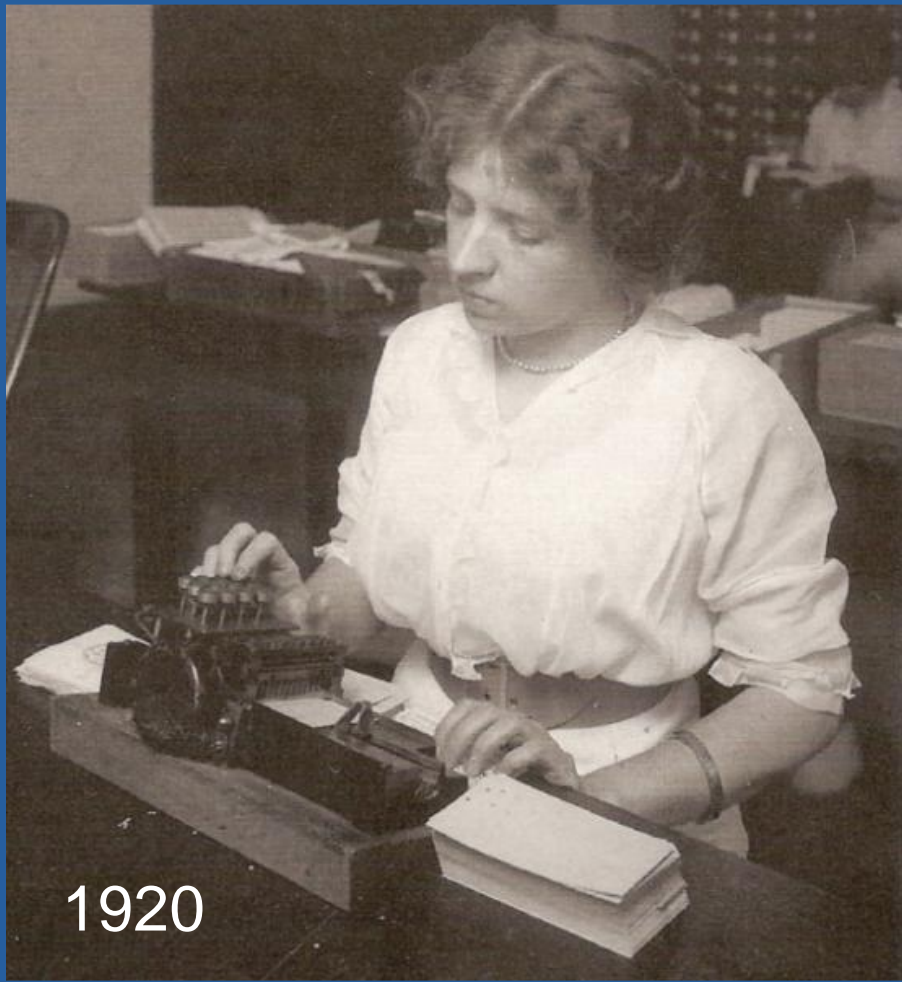
(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

(2) **make any publication whereby the data furnished by any particular establishment or individual under this title can be identified**; or

(3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports.

Existing Disclosure Avoidance Techniques

- Record-swapping
- Item imputation (for nonresponse)
- Whole household imputation (for nonresponse)
- Top- and bottom-coding



Images https://www.census.gov/history/www/sights_sounds/photos/1920_photos.php and <https://flic.kr/p/YiyyBE>

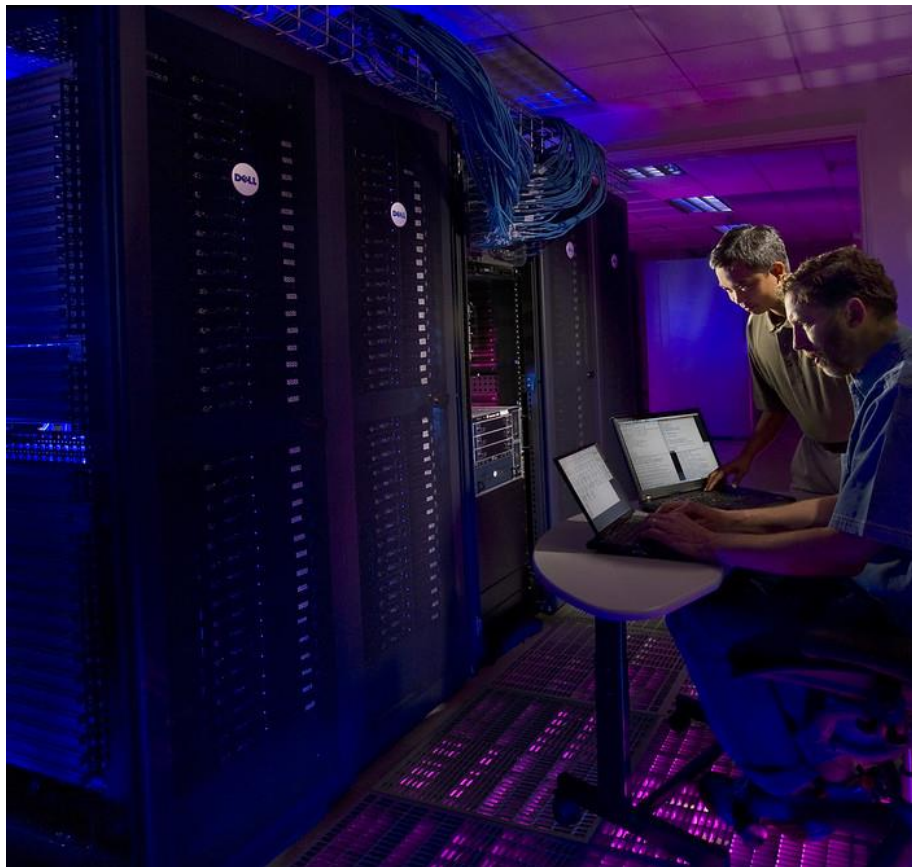
Reconstruction & Re-identification

Simulated reconstruction/re-ID “attack” on 2010 Census

Publication	Released Counts (incl 0s)
PL94-171 Redistricting	2,771,998,263
Balance of Summary File 1	2,806,899,669
SF2	2,093,683,376
PUMS	30,874,554
Est. published statistics	7,703,455,862
Statistics/person	25

Source: <https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf>

Simulated “attack” on 2010 Census



- Reconstructed location (census block), sex, age, ethnicity for 142 million people (46% of U.S. population).
 - 71% within 1 year of age.
- Matched 38% (of 142 mil) with information from commercial databases.
- Correctly re-ID ~17% of U.S. population in 2010.

Source <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/privacy-methods-2020-census.html> Image <https://flic.kr/p/fvdBcY>

Sea change in disclosure avoidance



Image Kodiak, AK <https://flic.kr/p/KQq3k9>



How does **Differential Privacy** Work?

Differential Privacy is not “new”

The screenshot displays the OnTheMap web application. The browser address bar shows the URL onthemap.ces.census.gov. The application header includes the title "OnTheMap" and links for "LEHD Home", "Help", and "and". The main interface is divided into a control panel on the left and a map area on the right. The control panel has tabs for "Start", "Base Map", and "Selection". It includes a "Welcome to OnTheMap!" section with instructions, a "Search" section with a search bar and a "Search All Names" dropdown, an "Import Geography" section with links for "Import from KML", "Import from SHP", and "Import from GPS", and a "Load .OTM File" section with a "Load" button. The map area shows a map of Alaska with Anchorage labeled. A scale bar at the bottom left of the map indicates 500 km and 200 mi. The footer of the application contains the text "PRB | INFORM. EMPOWER. ADVANCE." and a copyright notice for 2020 Population Reference Bureau.

OnTheMap

LEHD Home Help and

Start Base Map Selection

Save Load Feedback Previous Extent Hide Tabs

▼ Welcome to OnTheMap!

Start an analysis by using one of the tools below (Search, Import Geography, or Load .OTM file). Hover over the Help icons located throughout the application to see Help tips for using specific functionality. Sections in the control panel can be collapsed or opened by clicking the section title

2016 and 2017 data now available!
[Read New Data Notice \(08/29/2019\)](#)

▼ Search

Search

Search All Names

▼ Import Geography

[Import from KML](#)
[Import from SHP](#)
[Import from GPS](#)

▼ Load .OTM File

Click the "Load" button below to load a .OTM file.

Anchorage

500 km
200 mi

Privacy Policy | 2010 Census | Data Tools | Information

Source: U.S.Census Bureau, Center for Economic Studies

PRB | INFORM. EMPOWER. ADVANCE.

© 2020 Population Reference Bureau. All rights reserved. www.prb.org

Hypothetical example: “True” data

	American Indian or Alaska Native (AIAN)		
	AIAN Alone	AIAN in Combination	Not AIAN
Male	3	12	33
Female	4	17	31

Population = 100

Source: Example derived from David Van Riper, Minnesota Population Center, presentation to the Association of Public Data Users 2019.

Add noise to cross-tab

	American Indian or Alaska Native (AIAN)		
	AIAN Alone	AIAN in Combination	Not AIAN
Male	$3 - 1 = 2$	$12 + 0 = 12$	$33 + 1 = 34$
Female	$4 + 8 = 12$	$17 + 2 = 19$	$31 - 2 = 29$

Population now = 108

Pure differential privacy would end at this step, but...

Invariants (no noise added)

- Total population (for states, D.C. , Puerto Rico)
- Total housing units (census block level)
- Number of group quarters *facilities* by type (census block level)

For more information on invariants for 2020 Census, see:

<https://content.govdelivery.com/accounts/USCENSUS/bulletins/2ae5eda>

Enforce “invariants”

	American Indian or Alaska Native (AIAN)		
	AIAN Alone	AIAN in Combination	Not AIAN
Male	2	11	31
Female	11	18	27

Population = 100

Hypothetical (True) vs. Noise infused table

	American Indian or Alaska Native (AIAN)		
	AIAN Alone	AIAN in Combination	Not AIAN
Male	(3) 2	(12) 11	(33) 31
Female	(4) 11	(17) 18	(31) 27

Population = 100

Noise is independent of the size of any cell, but +/- start to balance out as you aggregate (larger geographies, larger groups).



How does this work nationwide?

Top Down

Step 1: National Data

Characteristic	Possible Values
Race (alone or in combination)	63
Ethnicity (Hispanic/Latino and not Hispanic/Latino)	2
Age (voting age, total population)	2
Occupancy status (occupied, vacant)	2
Group quarters (7 types)	7

~3,500 cells
(63 x 2 x 2 x 2 x 7)

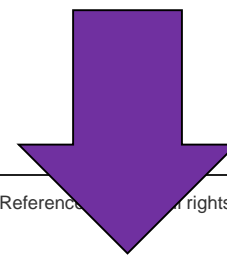
**“True”
National
Data**



**Noise
Infusion**

**National
Microdata**
(No geo IDs)

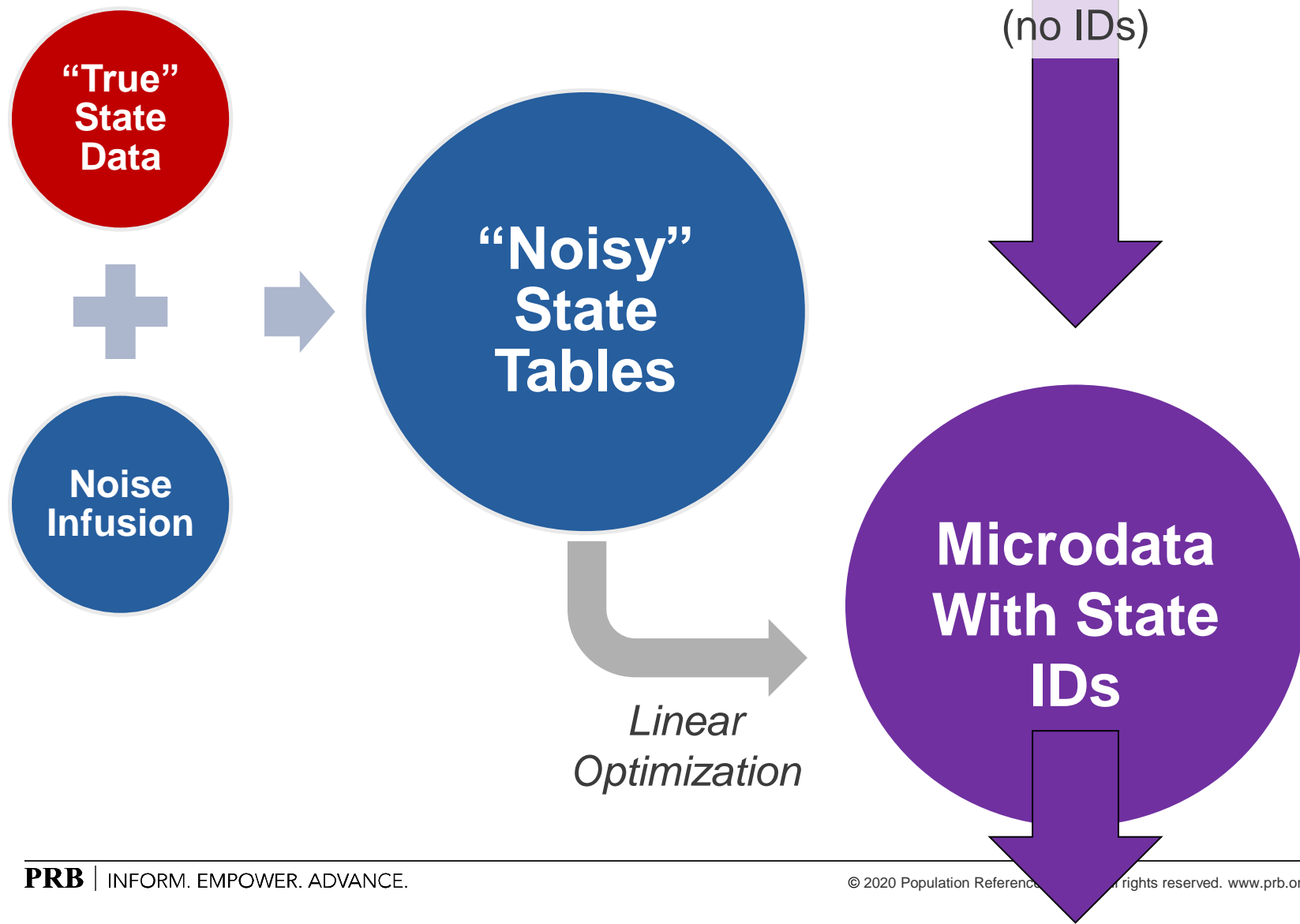
~330 million records



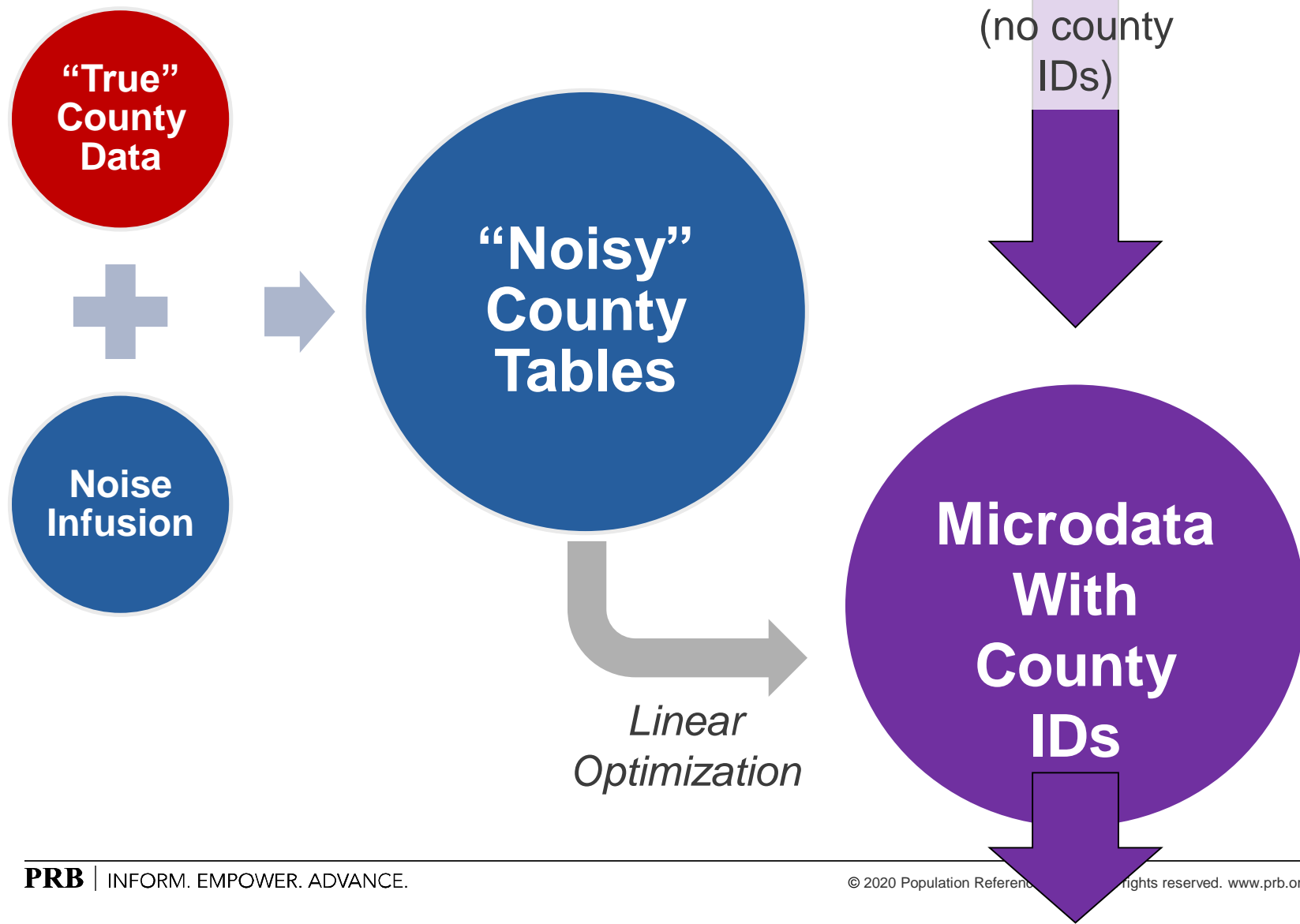
Source:

<https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf>

Step 2: State Data



Step 3: County Data



**Steps 4 & 5:
Repeat process
for all
geographies,
ending with
Census Blocks**

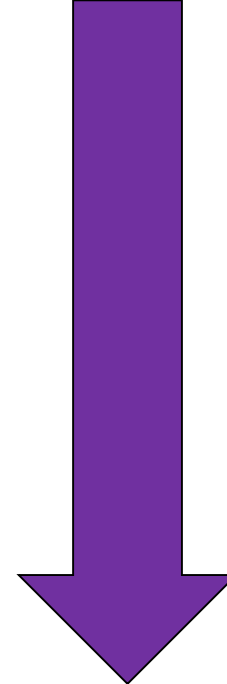
Repeat...
to Census
Tracts and
Census
Blocks

**Microdata
With State,
County,
Tract, Block
IDs**

Final Product: Noisy Microdata & Published Tables

Record-level characteristics retained from Step 1

Geographic IDs assigned so final microdata are “best fitting” to the “noisy” cross tabs for each geo level.



**Published
Tables**
drawn from noise-
infused microdata



Planned Data Products



Image <https://flic.kr/p/dTjhKs>

PRB | INFORM. EMPOWER. ADVANCE.

© 2020 Population Reference Bureau. All rights reserved. www.prb.org

Planned 2020 Census Data Products

- Apportionment (state population totals) *expected early 2021*
- Redistricting File (P.L. 94-171)
- Demographic and Housing Characteristics File (DHC)
 - Congressional District Demographic and Housing Characteristics File
- Demographic Profiles
- And more...

For latest information on products see

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products.html>

Redistricting File (P.L. 94-171)

- All redistricting data will be published at the census block level
- Subjects include:
 - Voting age
 - Race
 - Hispanic or Latino origin
 - Housing occupancy status
 - Group quarters population (by major type)

Source “2020 Census Data Products Crosswalk”

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products.html>

Redistricting File (P.L. 94-171)

Planned 2020 Public Law 94-171 Redistricting Data Summary File Tables

2010 Table	2020 Table	Title
H1	H1	Occupancy Status
P1	P1	Race
P2	P2	Hispanic or Latino, and Not Hispanic or Latino by Race
P3	P3	Race for the Population 18 Years and Over
P4	P4	Hispanic or Latino, and Not Hispanic or Latino by Race for the Population 18 Years and Over
N/A	P5	Group Quarters Population by Major Group Quarters Type

Source “2020 Census Data Products Crosswalk”

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products.html>

Redistricting File (P.L. 94-171)

Planned 2020 Public Law 94-171 Redistricting Data Summary File Tables

2010 Table	2020 Table	Title
H1	H1	Occupation
P1	P1	Population 18 Years and Over
P2	P2	Population 18 Years and Over by Hispanic or Latino by
P3	P3	Population 18 Years and Over
P4	P4	Population 18 Years and Over by Hispanic or Latino, and Not Hispanic or Latino by
N/A	N/A	Population 18 Years and Over
		Group Quarters Population by Major Group
		Quarters Type

**All the same details as 2010
Census PLUS group quarters
population detail**

Source “2020 Census Data Products Crosswalk”

<https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products.html>

Demographic and Housing Characteristics File (DHC)

- Subjects include:
 - Age
 - Sex
 - Race
 - Hispanic/Latino origin
 - Household type
 - Family type
 - Relationship to householder
 - Group quarters population
 - Housing occupancy
 - Housing tenure
- Some subjects provided for major OMB race/ethnicity groups
- Geographic level will vary by table

Source “2019-11-8: Status Update on the 2020 Census Data Products Plan”
<https://www2.census.gov/cac/nac/meetings/2019-11/devine-hollingsworth-status-update-2020-data-products-plan.pdf>

Demographic Profile

- Subjects include:
 - 5-year age groups
 - Sex
 - Race
 - Hispanic/Latino
 - Household type
 - Relationship to householder
 - Group quarters population
 - Housing occupancy
 - Housing tenure
- Geography: Place / minor civil division

Source “2019-11-8: Status Update on the 2020 Census Data Products Plan”
<https://www2.census.gov/cac/nac/meetings/2019-11/devine-hollingsworth-status-update-2020-data-products-plan.pdf>

Other products...

- Tables that require:
 - Complex person/household joins
 - Detailed race and Hispanic Origin groups
 - Data for American Indian and Alaska Native groups
- Microdata?

Early products will use TopDown. Later products may use other techniques as part of the Disclosure Avoidance System.

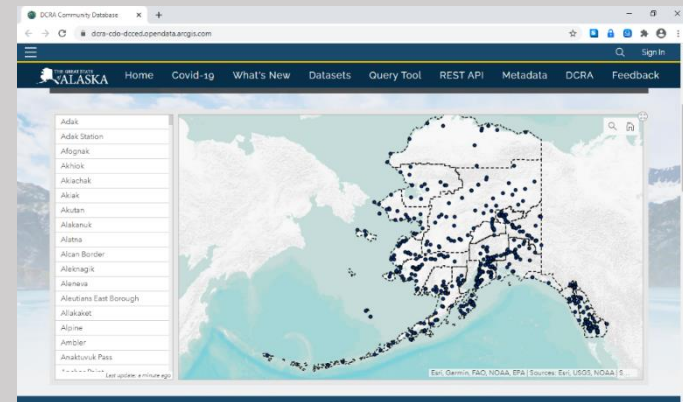
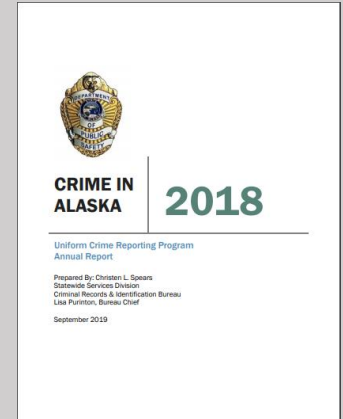
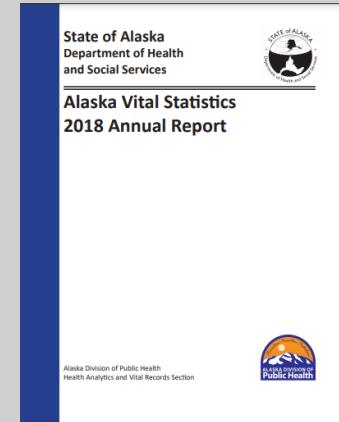
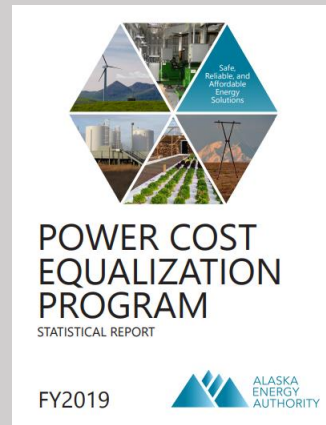
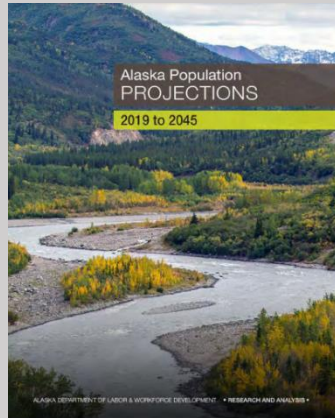
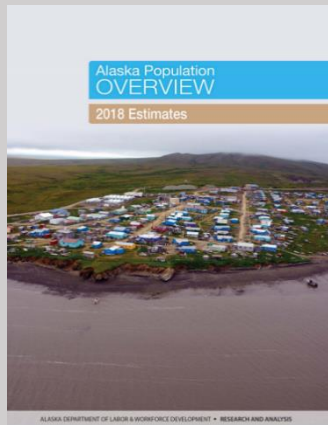
TBD...

- Level and distribution of noise
- Details of disclosure avoidance plans for later products

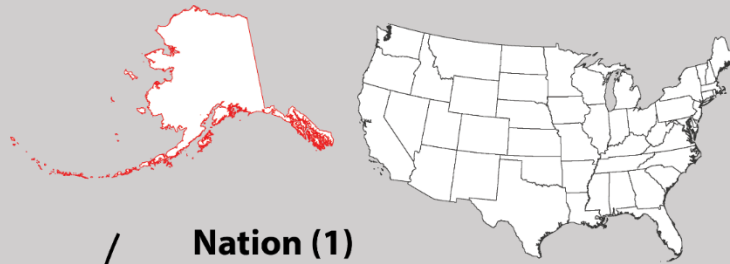


Image <https://www.census.gov/library/photos/tabulation-and-processing.html>

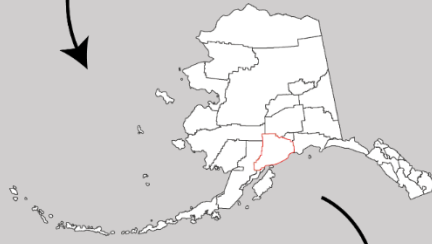
Decennial Census data is the foundation of nearly all Alaska Population Statistics



Census Geography breakdown



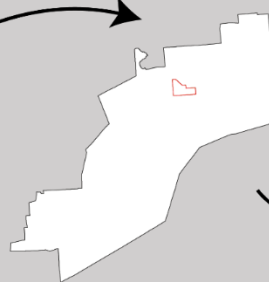
Nation (1)



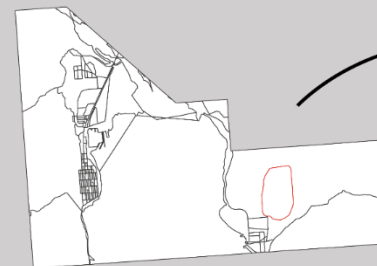
State (1)



**Borough/Census
Area (County) (29)**



Census Tract (167)



Block Group (534)

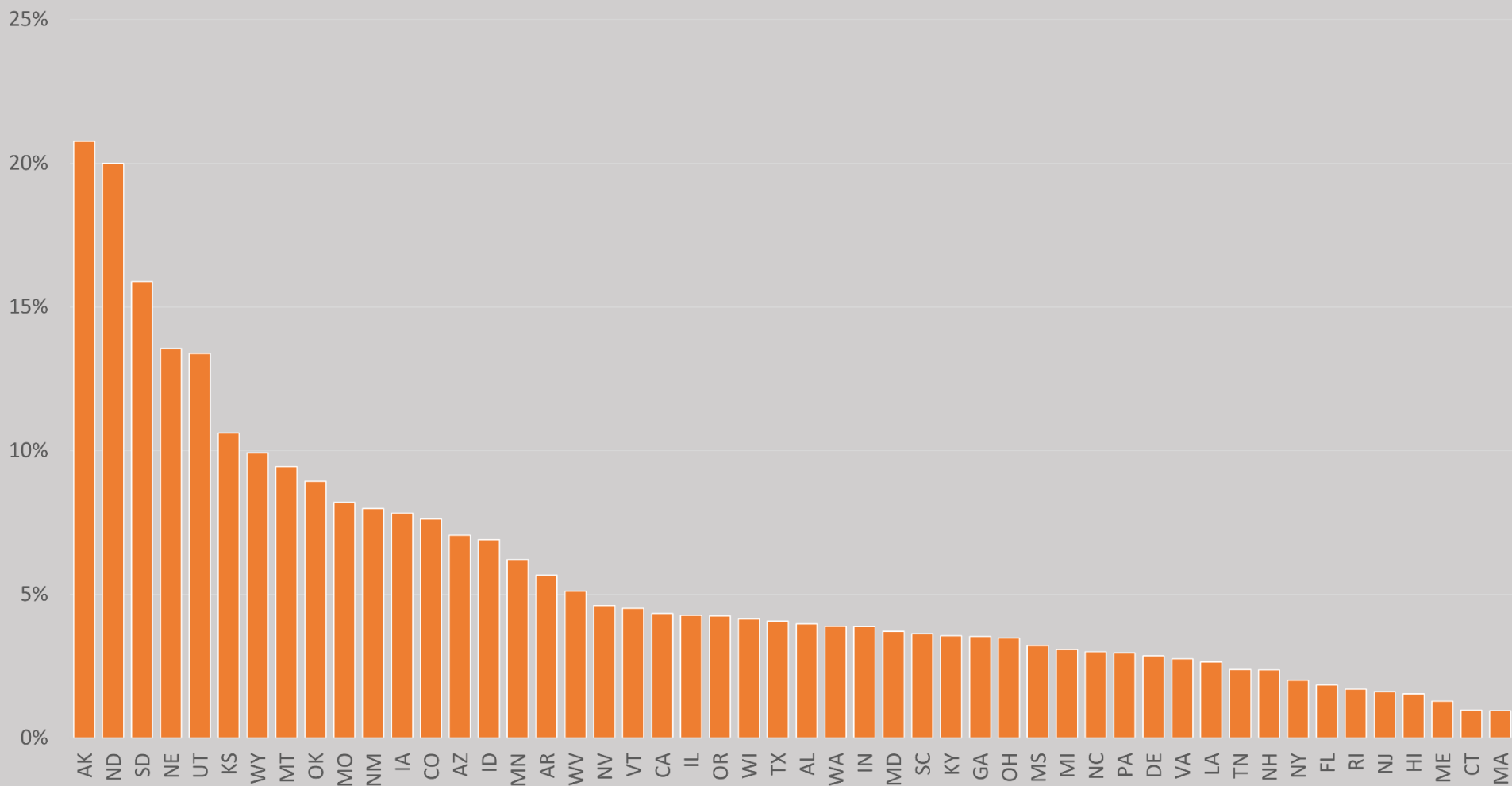


Block (45,292)

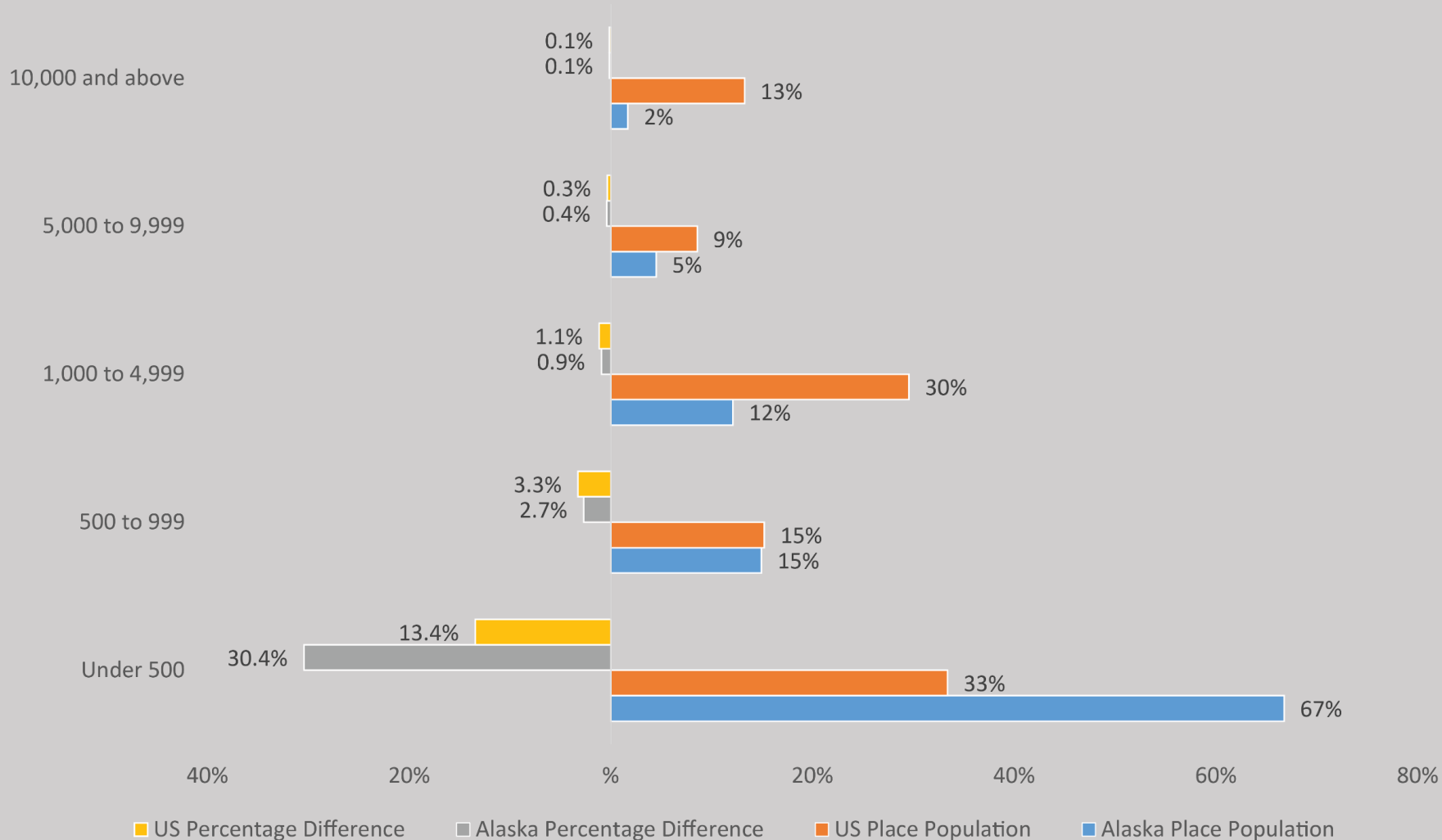
	Original Data	Privacy Budget	Privacy Protected Microdata File (PPMF)		
			Total Population		
	Average Population		MAE	MAPE (%)	Rank
Nation	308,745,538	0.2	invariant	invariant	-
State	710,231	0.16	invariant	invariant	-
County	24,491	0.16	5.9	0.2%	5
Tract	4,253	0.16	6.5	0.3%	25
Block Group	1,330	0.16	5.7	0.8%	7
Block	55	0.16	5.5	51.2%	-

Note: Average Populations below Nation are Alaska-only; Block population data only includes original populated blocks.

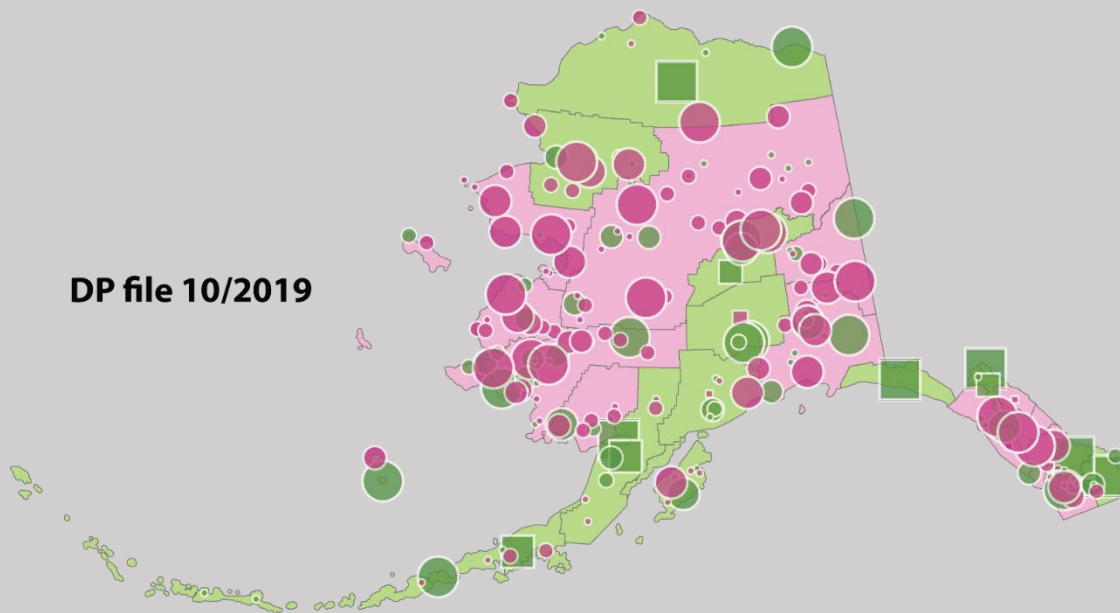
Average Percent Difference (MAPE) between PPMF and Original Count, 2010 Census Place-level



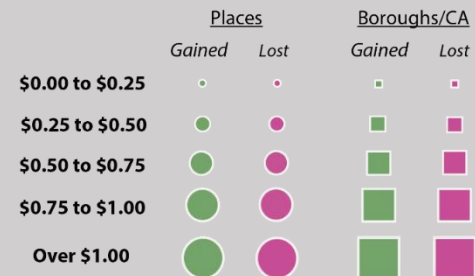
Population Distribution by Place and Average Percentage Difference from original count, 2010 Census PPMF



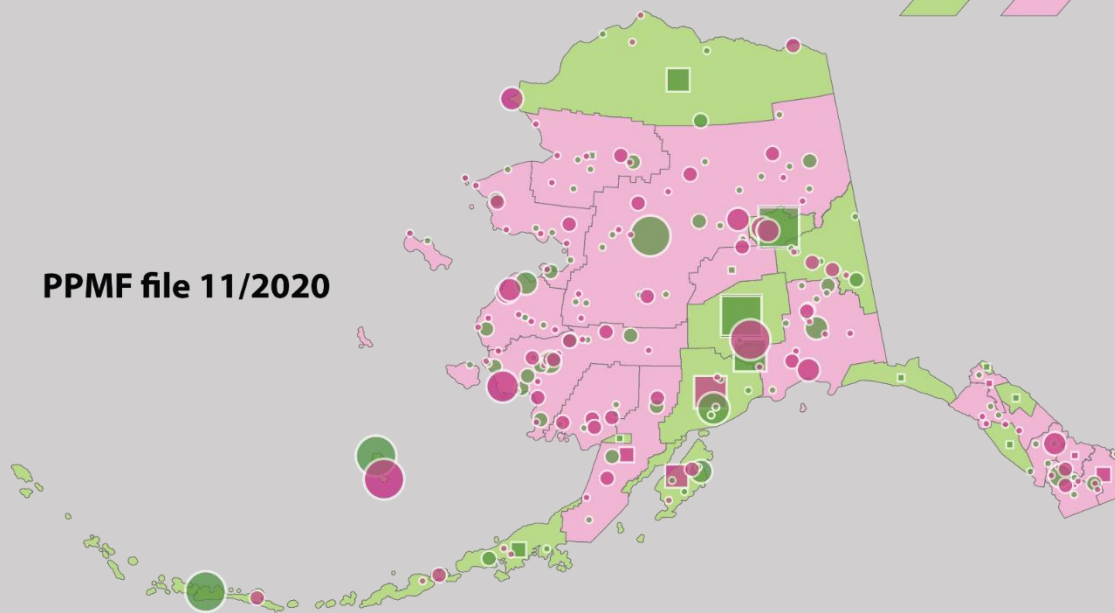
DP file 10/2019



Change in Per Capita Share per \$10,000



PPMF file 11/2020

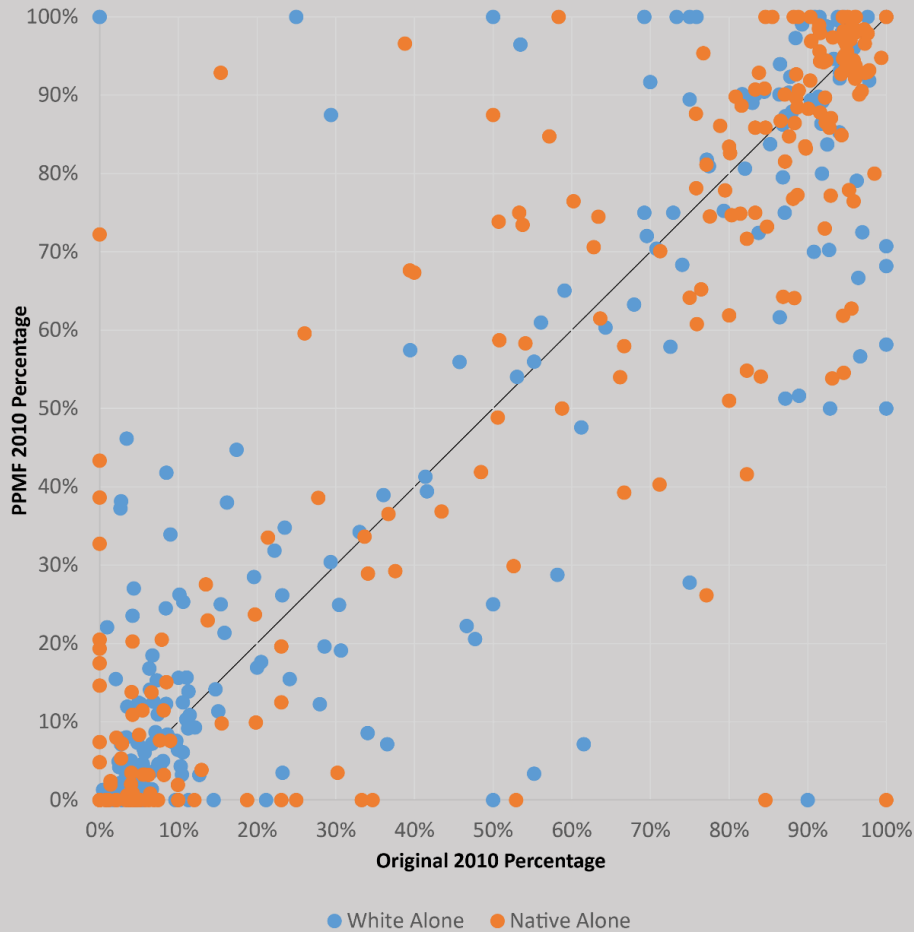


Borough/Census Area Total \$

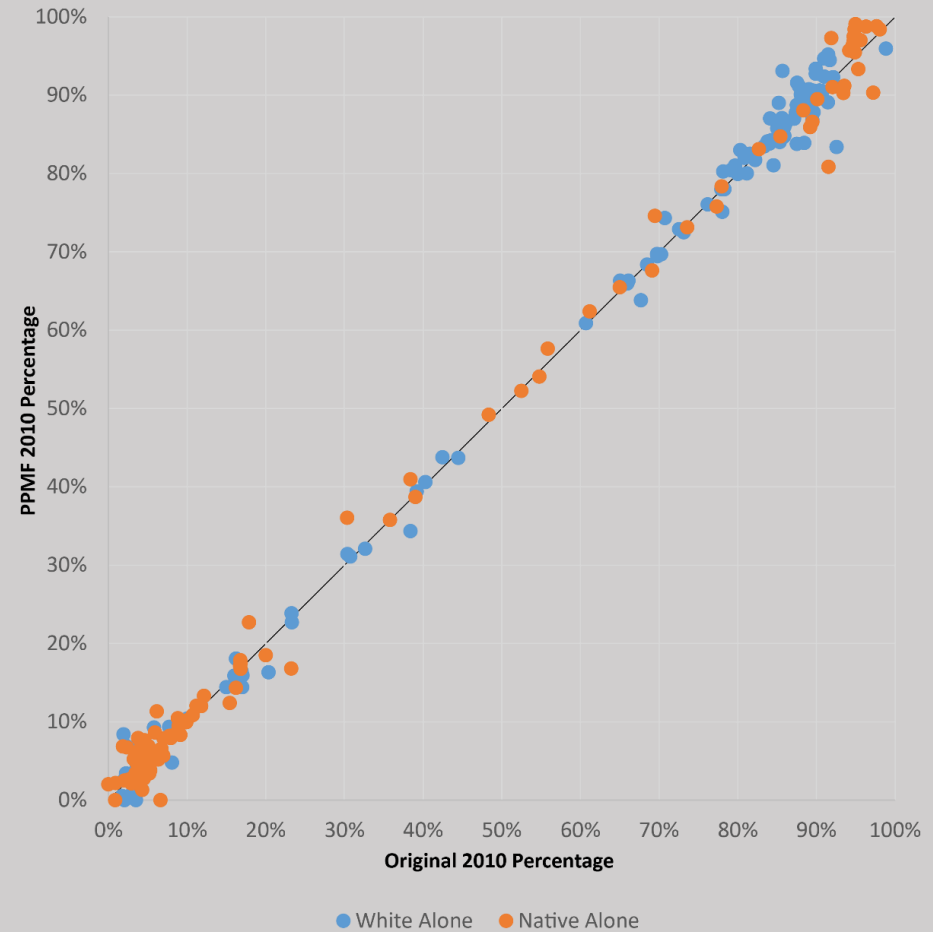


Comparison of Percentage White Alone and Native Alone by Place, PPMF vs Original Data 2010 Census

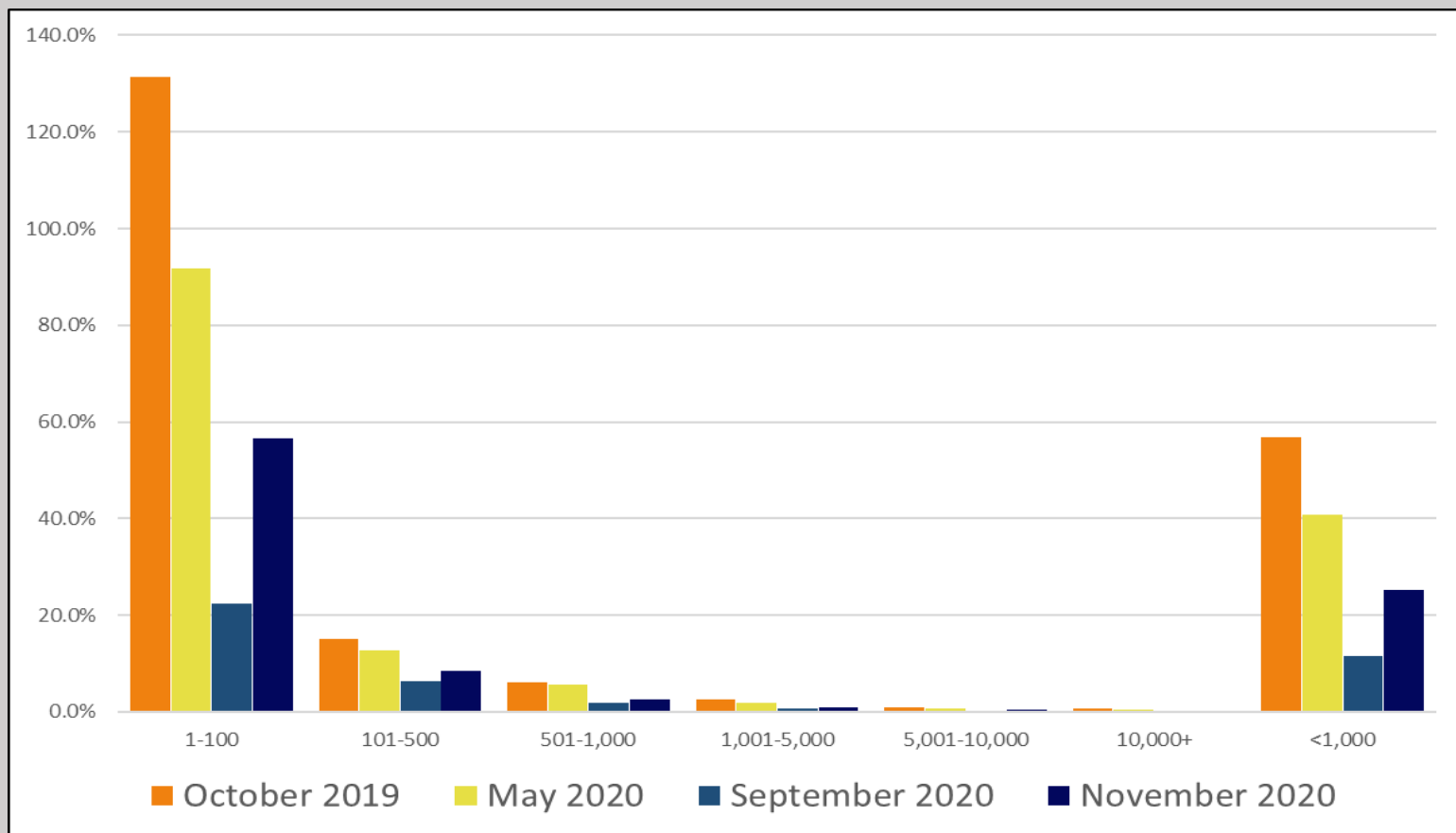
Places Under 500 People



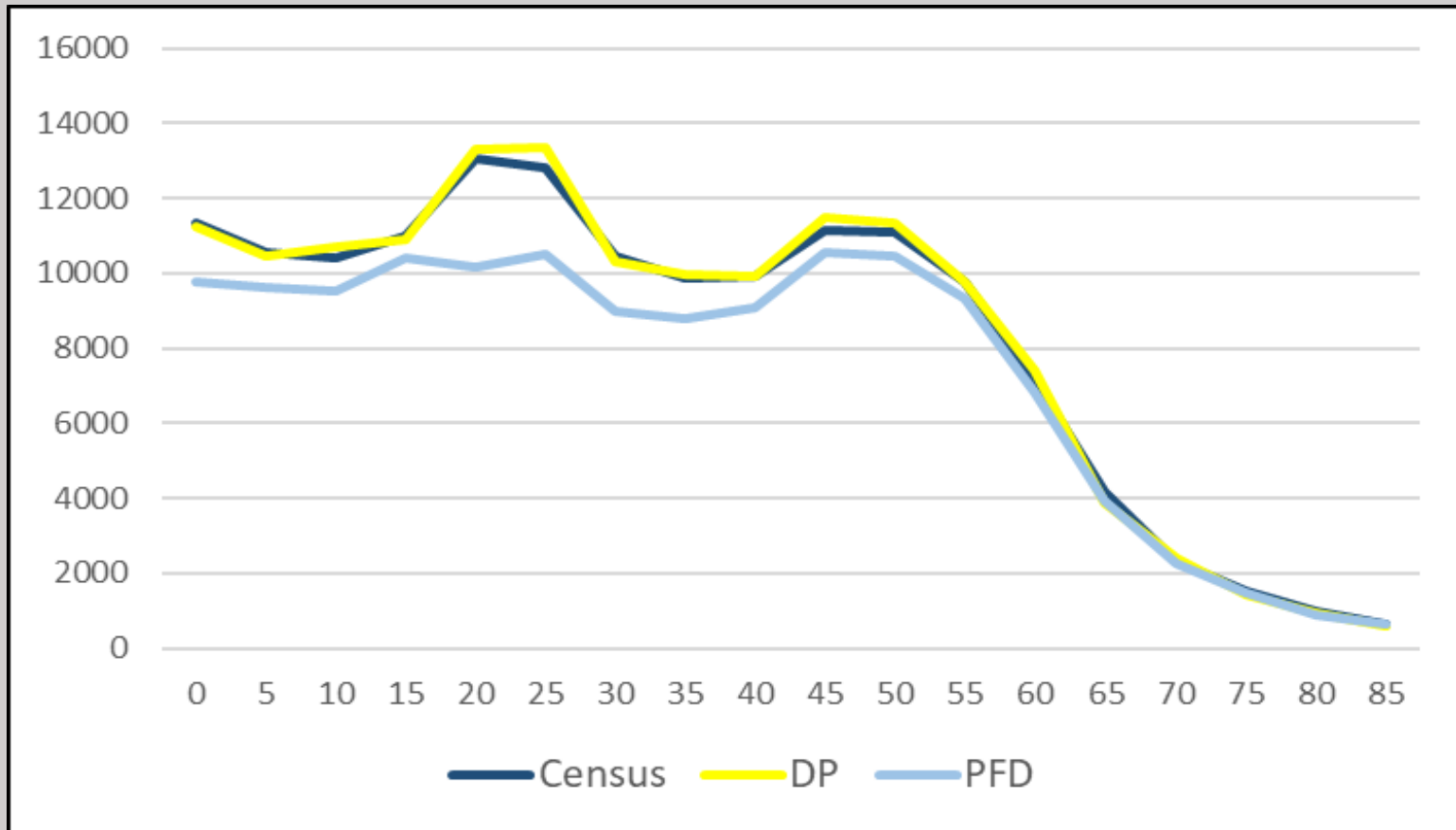
Places Over 500 People



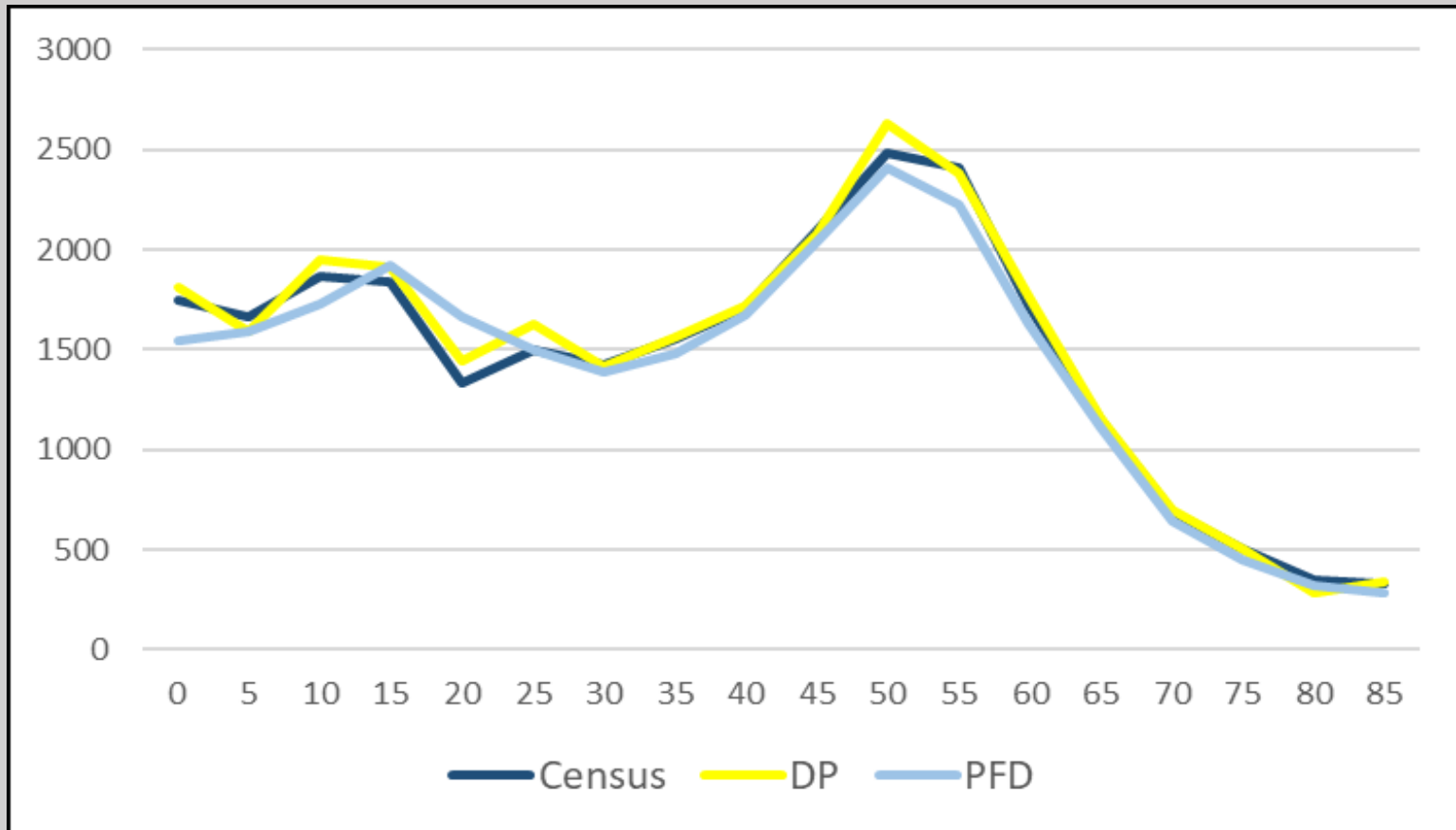
Mean Absolute Percent Difference by Population Size, Cities and CDPs



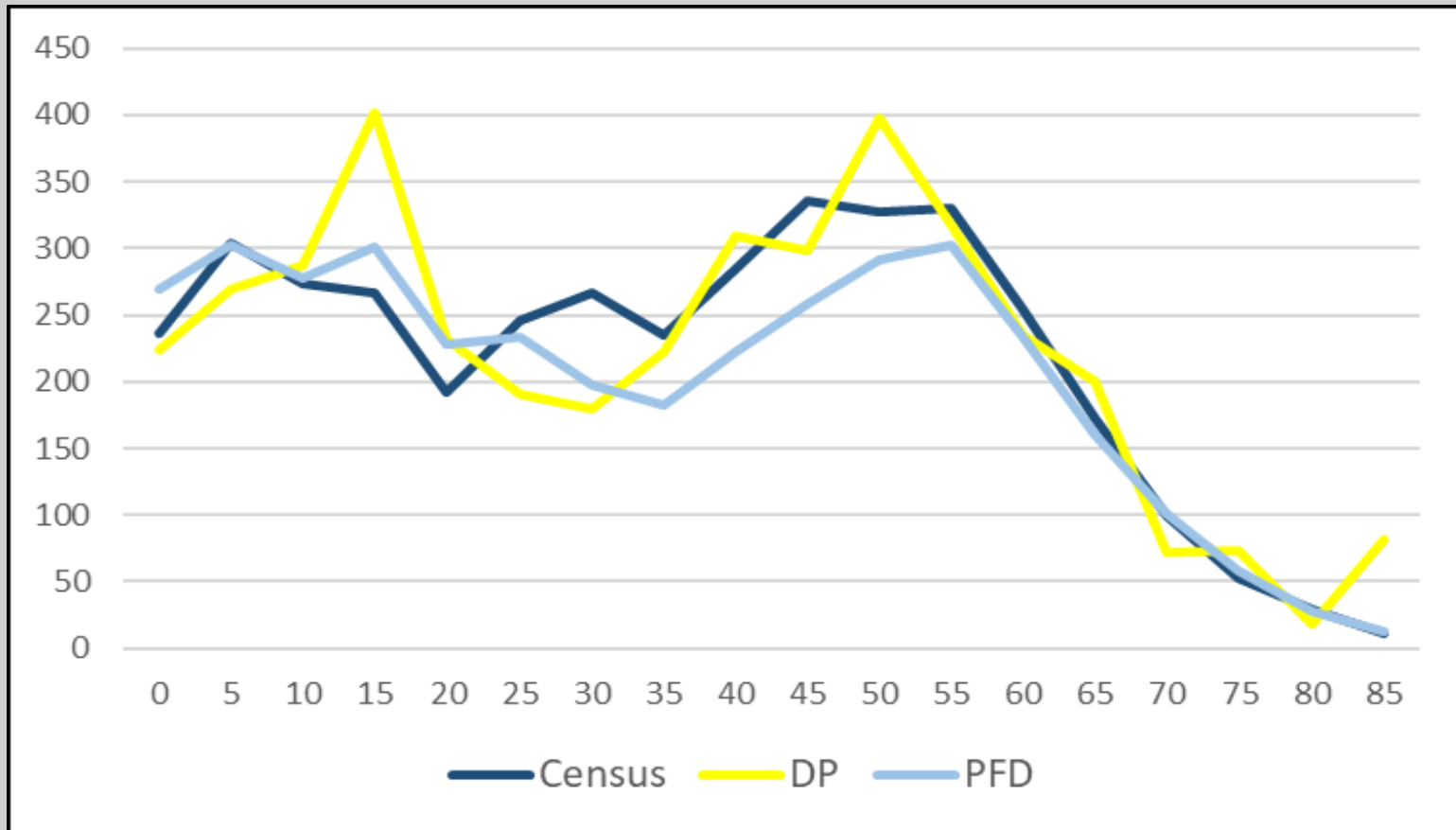
Anchorage Male by Age



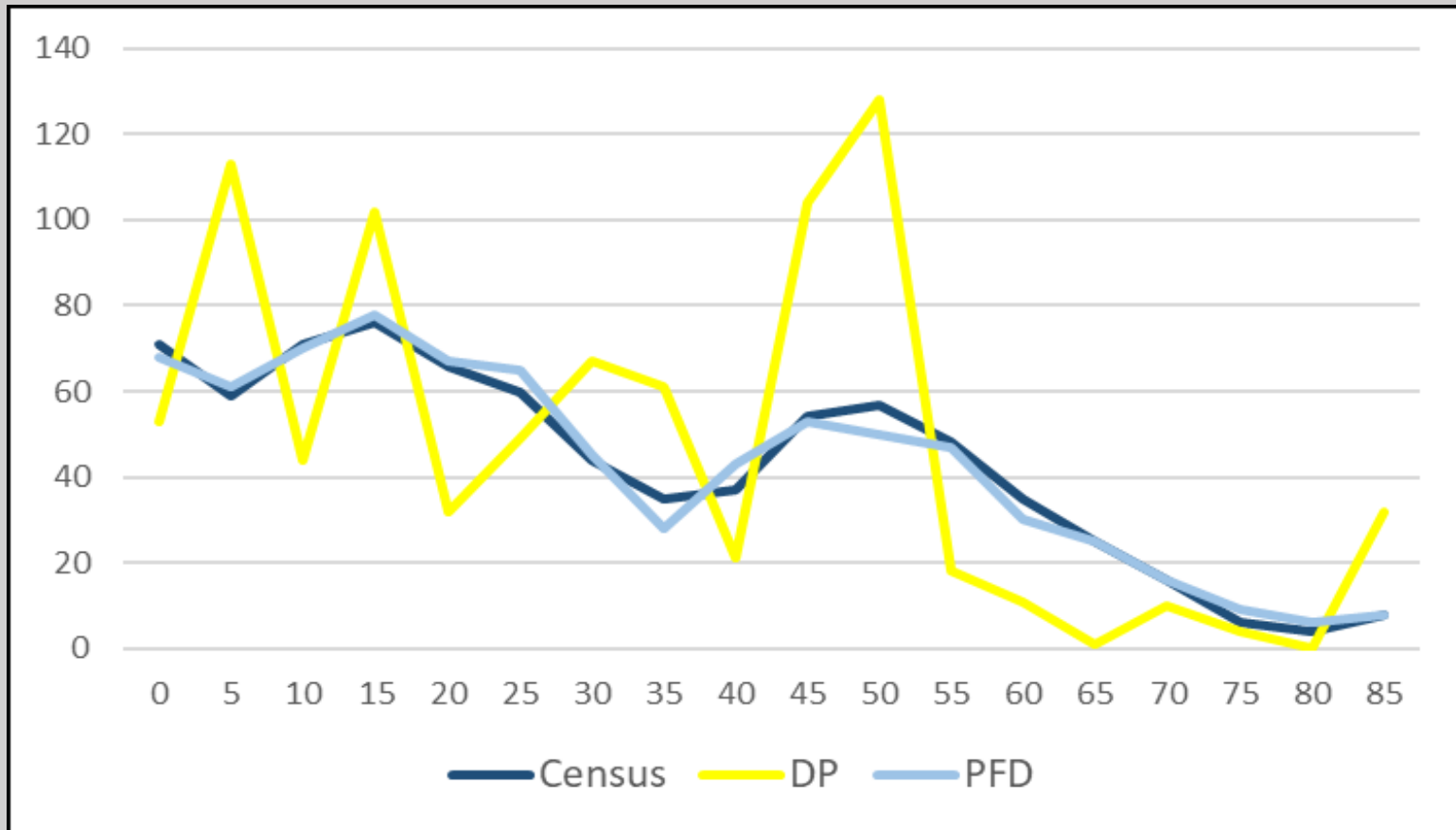
Kenai Female by Age



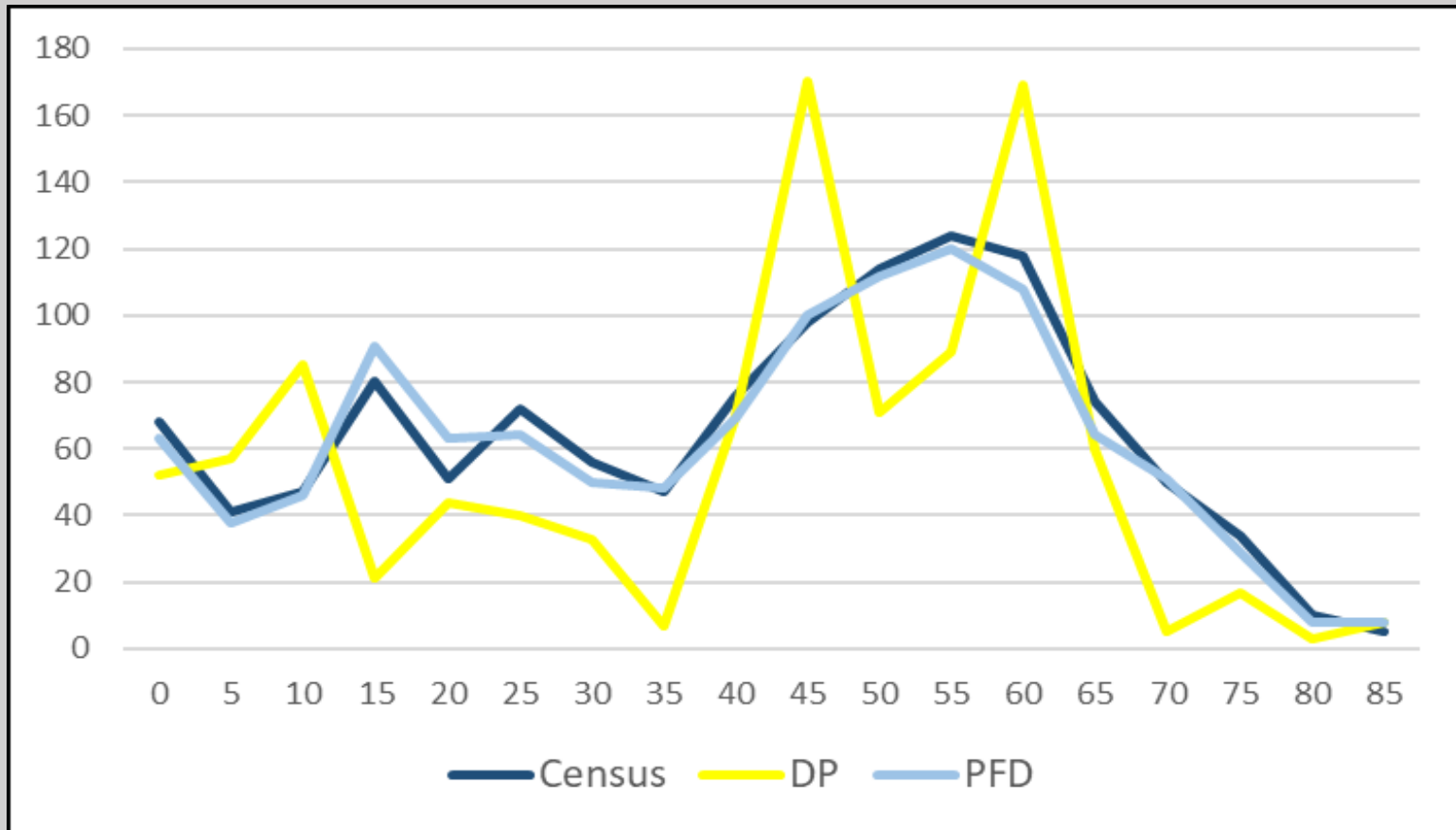
Southeast Fairbanks Male by Age



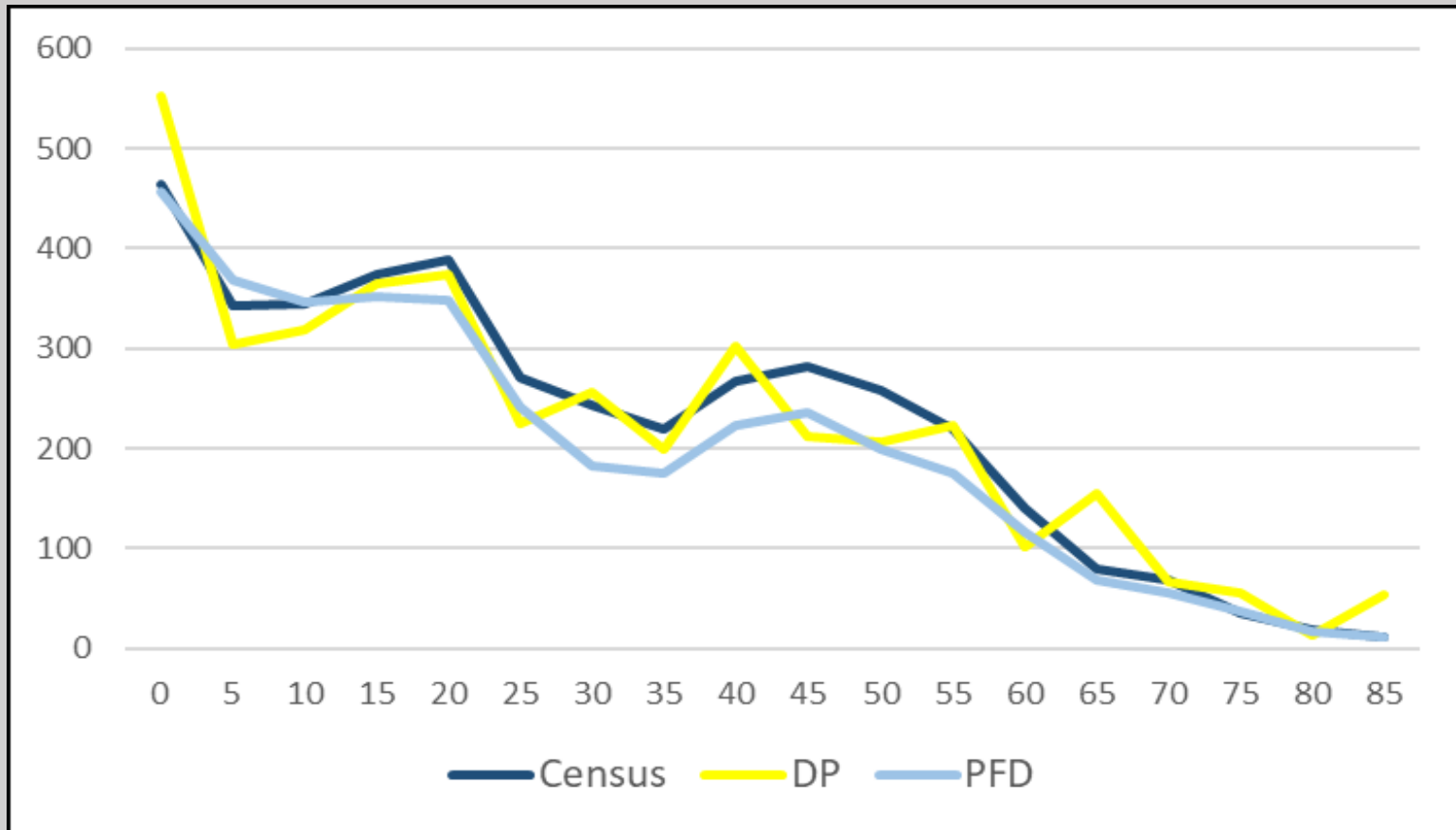
Lake and Peninsula Female by Age



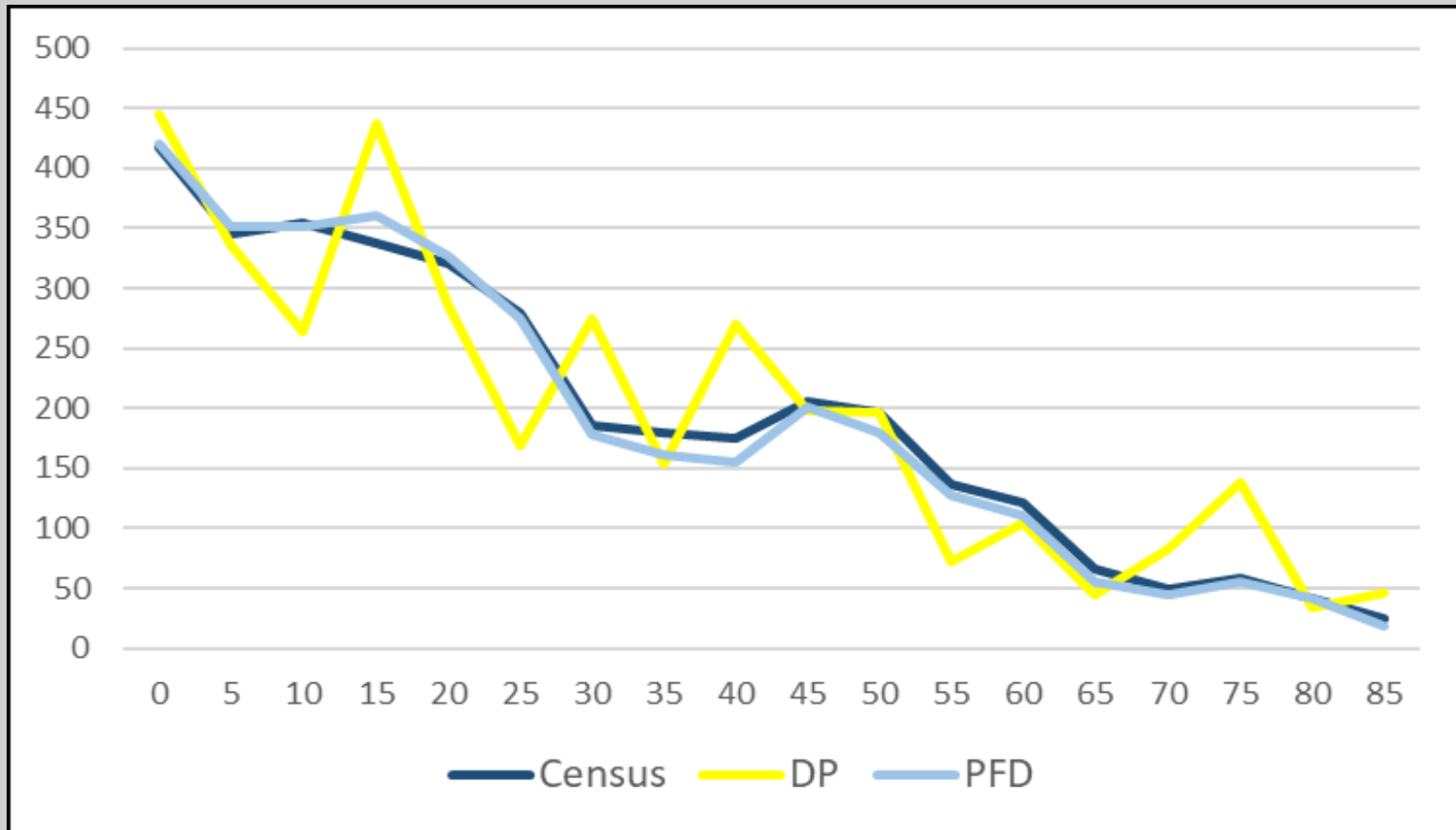
Hoonah-Angoon Male by Age



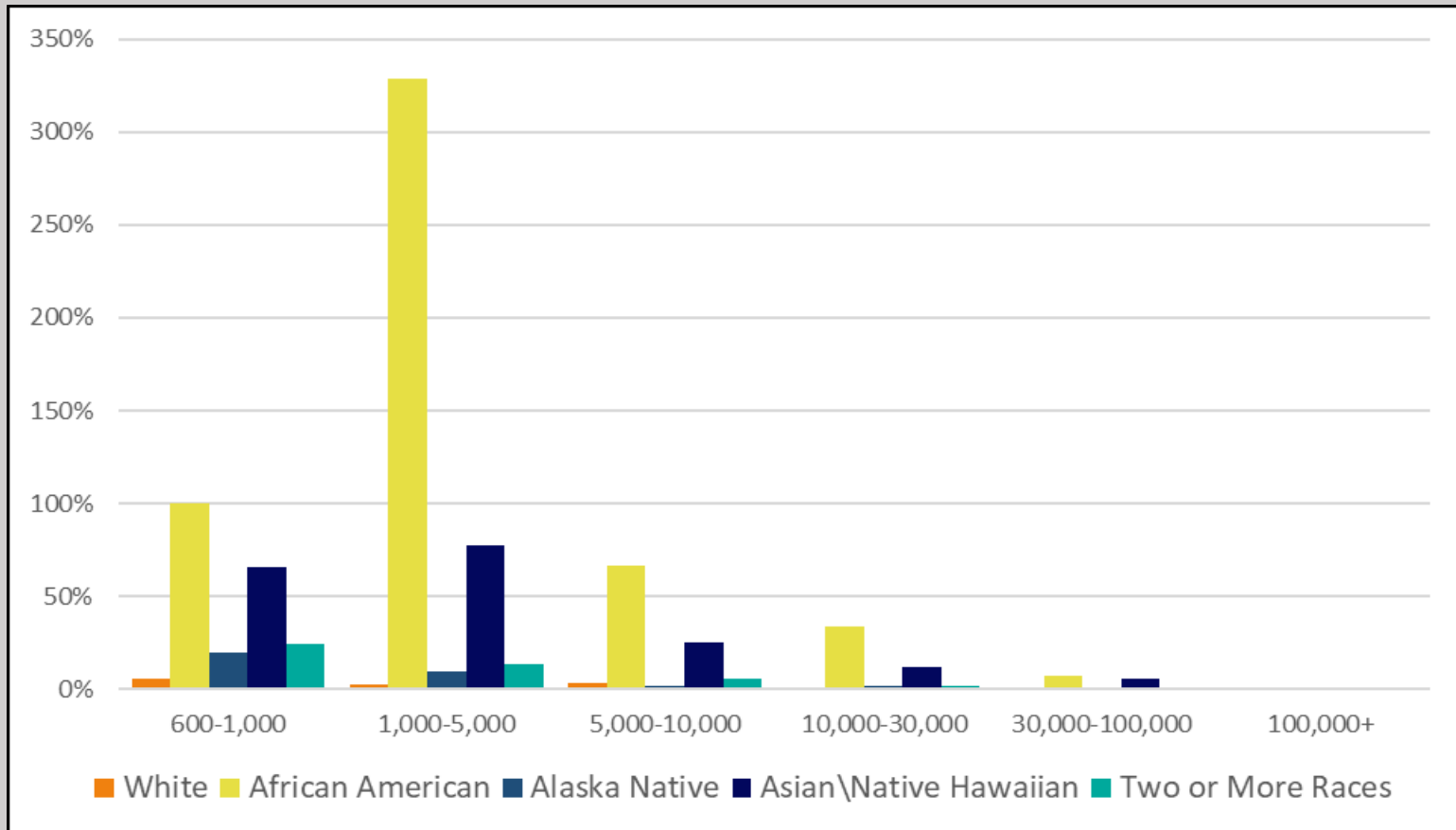
Northwest Arctic Male by Age



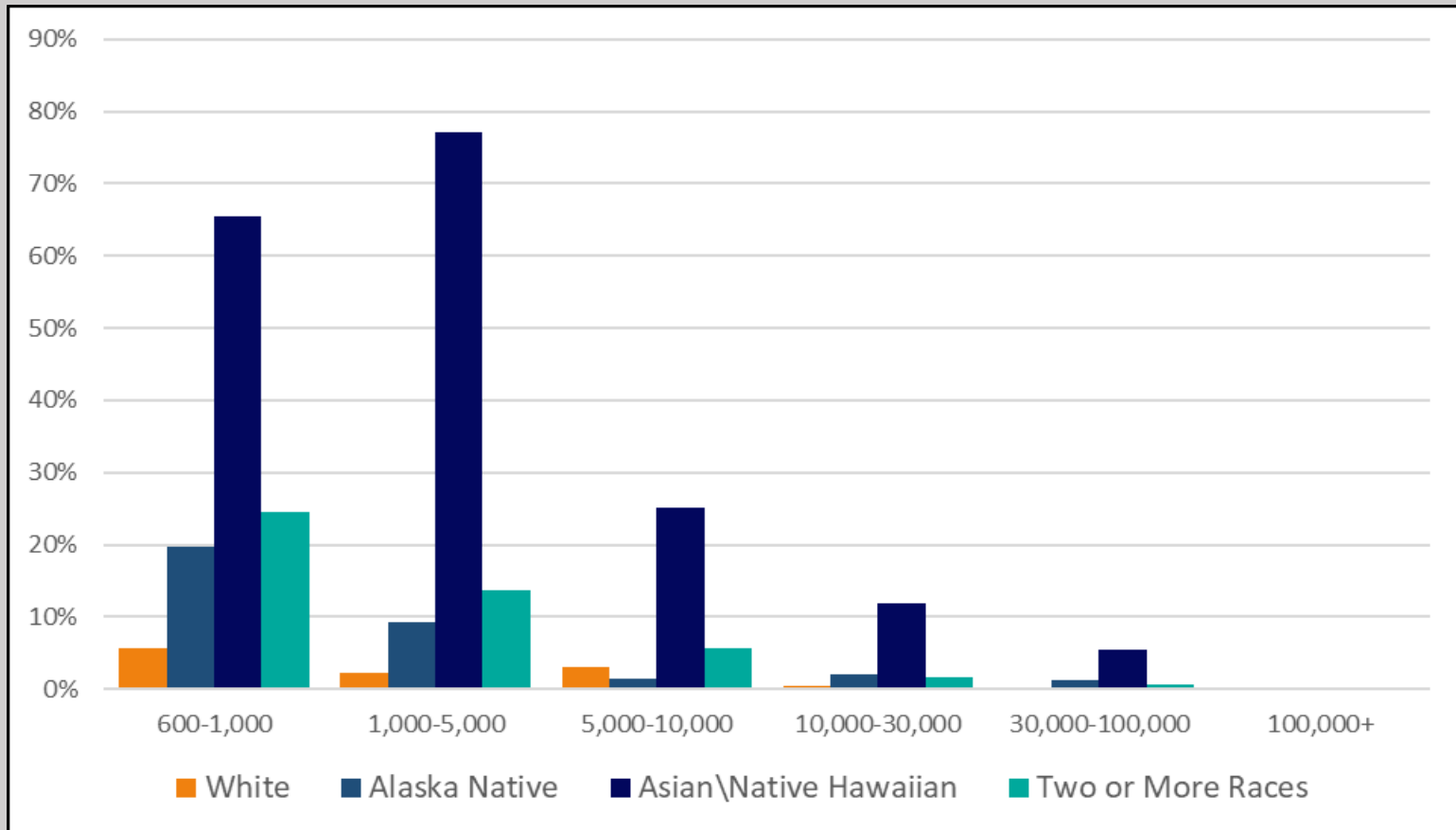
Northwest Arctic Female by Age



Mean Absolute Percent Difference by Population Size, BCA



Mean Absolute Percent Difference by Population Size, BCA



Thank you!

david.howell@alaska.gov

eric.sandberg@alaska.gov

Beth Jarosz on Twitter @DataGeekB



**ALASKA DEPARTMENT OF LABOR
& WORKFORCE DEVELOPMENT**